

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-225771
 (43)Date of publication of application : 22.08.1995

(51)Int.Cl. G06F 17/30

(21)Application number : 06-133105 (71)Applicant : FUJI XEROX CO LTD
 (22)Date of filing : 15.06.1994 (72)Inventor : HAYASHI SHOICHI

(30)Priority

Priority number : 05294211 Priority date : 30.10.1993 Priority country : JP

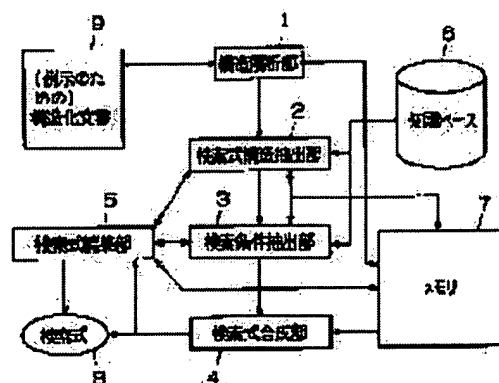
(54) RETRIEVAL EXPRESSION PREPARATION SUPPORT SYSTEM

(57)Abstract:

PURPOSE: To provide a retrieval expression

preparation support system which is capable of easily obtaining a right conditional expression for retrieval.

CONSTITUTION: A structured document 9 to be used for the illustration of a retrieval result is analyzed in a structure analysis part 1 and the structure is extended in a memory 7. A retrieval expression structure extraction part 2 successively takes out the routine structure patterns stored in a knowledge base 6 and finds out the partial structure pattern including the partial structure of the structured document illustrated by a user. A retrieval condition extraction part 3 fixes the value of the field to be used for a retrieval expression for each of the record included in the partial structure extracted in the retrieval expression structure extraction part 2. A retrieval expression synthesis part 4 synthesizes the expression of a form which is possible to be recognized as the retrieval expression. The synthesized retrieval expression is possible to be deformed in accordance with the edition instruction of the user and under such a condition that the illustrated partial structure is possible to be retrieved in a retrieval expression edition part 5.



LEGAL STATUS

[Date of request for examination] 14.06.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3168829

[Date of registration] 16.03.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-225771

(43)公開日 平成7年(1995)8月22日

(51)Int.Cl.⁶

G 0 6 F 17/30

識別記号

庁内整理番号

F I

技術表示箇所

9194-5L

G 0 6 F 15/ 403

3 3 0 A

審査請求 未請求 請求項の数4 O L (全 14 頁)

(21)出願番号 特願平6-133105

(22)出願日 平成6年(1994)6月15日

(31)優先権主張番号 特願平5-294211

(32)優先日 平5(1993)10月30日

(33)優先権主張国 日本 (J P)

(71)出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂三丁目3番5号

(72)発明者 林 祥一

神奈川県横浜市保土ヶ谷区神戸町134番地

横浜ビジネスパークイーストタワー 富

士ゼロックス株式会社内

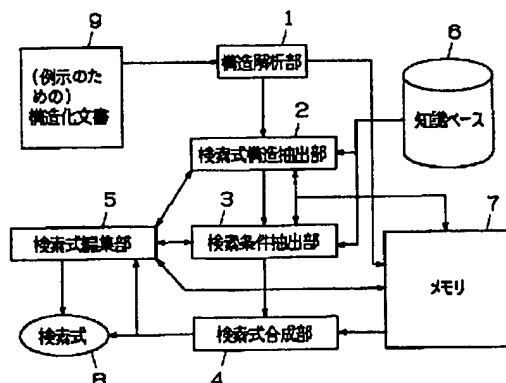
(74)代理人 弁理士 石井 康夫 (外1名)

(54)【発明の名称】 検索式作成支援システム

(57)【要約】

【目的】 検索のための正しい条件式を容易に得ることのできる検索式作成支援システムを提供する。

【構成】 検索結果の例示のために用いられる構造化文書9は、構造解析部1で解析され、その構造がメモリ7に展開される。検索式構造抽出部2は、知識ベース6の中に蓄積されている定型的な構造パターンを順次取り出し、ユーザにより例示された構造化文書の部分構造を含む部分構造パターンを見つけ出す。検索条件抽出部3は、検索式構造抽出部2で抽出された部分構造に含まれるレコードの各々について、検索式のために用いるフィールドの値を確定する。検索式合成部4は、検索式として認識可能な形式の表現を合成する。合成された検索式は、検索式編集部5で、ユーザの編集指示に従い、また、例示された部分構造を検索可能である条件のもとで、変形が可能である。



1

【特許請求の範囲】

【請求項1】 構造化文書に対して検索処理を行なう際に用いる検索式の作成を支援する検索式作成支援システムにおいて、前記構造化文書が記憶される記憶手段と、検索結果の1つとしてユーザから例示された前記構造化文書の部分構造に基づき前記例示された部分構造を含む部分構造を抽出する構造抽出手段と、該構造抽出手段により抽出された部分構造から検索式を合成する検索式合成手段を有することを特徴とする検索式作成支援システム。

【請求項2】 ユーザからの編集操作指示に従い、前記検索式合成手段により合成された検索式に対して変形を行なう際に、変形された検索式が合成の際に用いた部分構造を含むことを条件に検索式の変形処理を実行する検索式編集手段を有することを特徴とする請求項1に記載の検索式作成支援システム。

【請求項3】 構造化文書に対して検索処理を行なう際に用いる検索式の作成を支援する検索式作成支援システムにおいて、検索式に従い検索を行なう検索手段と、該検索手段による検索結果に基づきユーザが検索結果に対する誤りの指摘を行なう入力手段と、該入力手段から入力される誤りの指摘に従い指摘された部分を含む部分構造と検索式の構造に基づき検索式を修正する検索式修正処理手段を有することを特徴とする検索式作成支援システム。

【請求項4】 誤りが指摘された検索式の構造に含まれている各枝と、その枝の組み合わせに関する情報を順次記憶するデータベースを有し、前記検索式修正処理手段は、検索式を修正する際に前記データベースを参照し、前記データベースに記憶されている枝の組み合わせを含む検索式の修正候補以外の修正候補に基づいて検索式の修正を行なうことを特徴とする請求項3に記載の検索式作成支援システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、構造化文書を処理する構造化文書処理装置において、構造化文書に対して検索処理を行なう際に、検索に用いる検索式の作成を支援する検索式作成支援システムに関するものである。

【0002】

【従来の技術】 従来の構造化文書処理装置には、構造化文書における文脈、すなわち、階層構造に関する情報を含むパターンを検索条件として与え、検索を行なうことができるものがある。また、このような検索式の作成を支援する装置も考えられている。この装置を用いることによって、ユーザは条件式をグラフ表現によって入力し、それを編集することによって、検索条件を入力及び変更することができる。しかし、従来の構造化文書処理装置においては、ユーザ自身がはじめから検索条件を考えて入力しなければならない。そのためには、構造化文

2

書の内部表現がどのようなになっているかをユーザが正確に知っている必要がある。

【0003】 従来の検索式の作成をグラフ表現によって支援する装置では、入力装置が認識可能な形式を分かりやすく表現するもので、正確な文法に関する知識がないユーザにも文法的に正しい式を作成できるようにするものである。しかし、このような装置でも、構造化文書の内部表現に関する正確な知識をユーザに要求する点に変わりはない。

10 【0004】 また、従来の構造化文書処理装置は、正しい意味を持った式を作成することに関しては、何も支援するものではない。このため、検索が非常に有用な機能であるにもかかわらず、誰もが容易に利用できる機能となっていない。

【0005】

【発明が解決しようとする課題】 本発明は、上述した事情に鑑みてなされたもので、検索のための正しい条件式を容易に得ることのできる検索式作成支援システムを提供することを目的とするものである。

20 【0006】

【課題を解決するための手段】 本発明は、請求項1に記載の発明においては、構造化文書に対して検索処理を行なう際に用いる検索式の作成を支援する検索式作成支援システムにおいて、前記構造化文書が記憶される記憶手段と、検索結果の1つとしてユーザから例示された前記構造化文書の部分構造に基づき前記例示された部分構造を含む部分構造を抽出する構造抽出手段と、該構造抽出手段により抽出された部分構造から検索式を合成する検索式合成手段を有することを特徴とするものである。

30 【0007】 また、請求項2に記載の発明においては、ユーザからの編集操作指示に従い、前記検索式合成手段により合成された検索式に対して変形を行なう際に、変形された後の検索式が合成の際に用いた構造化文書の例示された部分構造を含むことを条件に検索式の変形処理を実行する検索式編集手段を有することを特徴とするものである。

40 【0008】 請求項3に記載の発明においては、構造化文書に対して検索処理を行なう際に用いる検索式の作成を支援する検索式作成支援システムにおいて、検索式に従い検索を行なう検索手段と、該検索手段による検索結果に基づきユーザが検索結果に対する誤りの指摘を行なう入力手段と、該入力手段から入力される誤りの指摘に従い指摘された部分を含む部分構造と検索式の構造に基づき検索式を修正する検索式修正処理手段を有することを特徴とするものである。

50 【0009】 請求項4に記載の発明においては、請求項3に記載の検索式作成支援システムにおいて、誤りが指摘された検索式の構造に含まれている各枝と、その枝の組み合わせに関する情報を順次記憶するデータベースを有し、前記検索式修正処理手段は、検索式を修正する際

3

に前記データベースを参照し、前記データベースに記憶されている枝の組み合わせを含む検索式の修正候補以外の修正候補に基づいて検索式の修正を行なうことを特徴とするものである。

【0010】

【作用】請求項1に記載の発明によれば、ユーザに検索結果の1例を例示させ、例示によって得られた構造化文書の部分構造を含むような部分構造を構造抽出手段で抽出し、その抽出した構造から検索式合成手段により検索式を自動合成することにより、文法的に正しい検索式を例示という簡単な操作により得ることができる。

【0011】また、合成された検索式の条件を強め、あるいは、ゆるめて検索したい場合がある。そのため、請求項2に記載の発明によれば、検索式編集手段によって合成された検索式の変形を行なうことができる。しかし、自由な変形を行なうと、例示した検索結果が含まれないという不合理な検索式が作成されてしまう場合がある。検索式編集手段では、変形された後の検索式が、合成の際に用いた部分構造を含むことを条件に、検索式の変形処理を実行する。これにより、ユーザは、合成の際に例示された部分構造を常に検索可能な、常に正しい検索式を得ることができる。

【0012】請求項3に記載の発明によれば、意図したものではない検索結果が得られた場合に、検索式側ではなく、検索結果の方の誤っている部分がどこであるのかをユーザが入力手段から例示することにより、検索式修正処理手段は、例示された部分を含む部分構造と検索式の構造に基づき、指摘された誤りをたどるように検索式を自動修正する。検索式修正処理手段は、指摘される誤りの情報が不完全、すなわち、全ての誤りが指摘されるのではなく、誤っている部分の一部の例示のみであっても、順次検索式を修正して行くことができる。このとき、誤りに関する指摘の数の増加に伴って、その全ての誤りを解決するような、より正しい検索式が得られるように修正することが可能である。

【0013】請求項4に記載の発明によれば、誤りが指摘された検索式の構造に含まれている枝の組み合わせに関する情報は、順次、データベースに記憶され、検索式修正処理手段が検索式を修正する際には、データベースに記憶されている情報、すなわち、誤りが指摘された検索式への修正を行なわないようにすることができる。これにより、過去の履歴を学習した一種の学習効果を得ることができ、その時点までに判明している同じ誤りを繰り返さないで、検索式の修正を行なうことができる。

【0014】

【実施例】図1は、本発明の検索式作成支援システムの第1の実施例の構成図である。図中、1は構造解析部、2は検索式構造抽出部、3は検索条件抽出部、4は検索式合成部、5は検索式編集部、6は知識ベース、7はメモリ、8は検索式、9は構造化文書である。

4

【0015】構造解析部1は、検索結果の例示のために用いられる構造化文書9を解析し、その構造をメモリ7に展開する。このとき、メモリ7には、例示に用いられる構造化文書9の階層構造が、親子関係と兄弟関係にあるものの間の双方向にポインタを持つレコードの木構造として記憶される。検索式構造抽出部2は、知識ベース6の中に蓄積されている定型的な構造パターンを順次取り出し、ユーザにより例示された構造化文書の部分構造を含む部分構造に一致するものを見つけ出す。そして、見つけ出した定型的な構造パターンに一致するレコード全てに対して、マークを付与する。検索条件抽出部3は、検索式構造抽出部2で抽出された部分構造に含まれるレコードの各々について、検索式のために用いるフィールドの値を確定する。検索式合成部4は、マークが付与された部分構造のレコードを順次取り出し、検索式として認識可能な形式の表現を合成する。検索式編集部5は、ユーザからの編集指示に従い、検索式合成部4により合成された検索式に対して変形を行ない、新たな検索式を生成する。このとき、それまでに指示されている部分構造を検索可能であるという正しさが失われない場合10に限り、検索式の変形を行なう。知識ベース6には、熟練したユーザが検索のために用いる文脈(部分構造)を定型化して格納しておく。

【0016】図2は、本発明の検索式作成支援システムの第1の実施例における検索式生成の動作を示すフローチャートである。S11において、ユーザは、例えば、構造化文書エディタ等を用いて、検索したい部分を例示する。例示は、例えば、構造化文書エディタに表示されているレコードを、例えば、ポインティングデバイス等で指示入力すればよい。この例示により、指示されたレコードの識別子を得ることができる。その識別子が一致するレコードを、メモリ7に展開されている例示のための構造化文書から見つけだし、メモリ7内の構造化文書のレコードにマークを付与する。

【0017】次に、S12、S13において、ユーザから例示された部分構造、すなわち、メモリ7内の構造化文書のマークを付与されたレコードを包含する部分構造を、知識ベース6を用いて抽出する。知識ベース6には、熟練したユーザが検索のために用いる文脈を部分構造として格納しており、ユーザから例示された部分構造を検索するために、熟練したユーザが検索条件として用いるであろう部分構造を蓄積してある。そのため、未熟なユーザからの指示であっても、熟練したユーザが用いるのと同程度の検索条件が得られる。抽出される部分構造は、ユーザから例示された部分構造を含む知識ベース6に格納されている構造の一つと同形の部分構造である。抽出された定型的な部分構造パターンに一致する、メモリ7内の構造化文書のレコード全てに対して、マークを付与する。マークを付与されたレコードは、メモリ7に展開されている文書の構造の部分構造(部分木)と

5

なる。そして、この部分構造が検索式の持つべき構造となっている。

【0018】S14において、抽出された部分構造に対応する検索式の条件を求める。S13までのステップで検索式の持つべき構造が決定されるが、各レコードの属性、例えば、図や表題、段落などの条件も検索の際に必要となる。そのため、メモリ7内の構造化文書の抽出された部分構造、すなわち、マークの付与されたレコードの属性を参照し、検索のための条件を求める。

【0019】S15において、検索のための部分構造及び条件に基づき、検索式合成部4において、検索式を生成し、S16において、生成した検索式を出力する。ユーザは、出力された検索式を用いて検索処理を行なえばよい。

【0020】上述の構成では、検索式構造抽出部2で定型的な構造パターンを抽出する際に、知識ベース6に保持しているルールを用いているが、知識ベース6内に保持されているルールには、ユーザの例示した部分構造を含むものがない場合には、デフォルトのルールにより対応する部分構造を設定したり、あるいは、推論により、類似の構造を抽出し、ユーザに提示するなどの手段を設けることもできる。また、知識ベース6を用いず、ユーザの例示した部分構造に対応する部分構造を返す手続きとして、プログラムの中に組み込むような構成をとることもできる。

【0021】図3は、構造化文書の一例の説明図、図4は、構造化文書の構造化された内部表現の一例の説明図である。図3に示した文書には、「序論」、「本論」等の表題や、それぞれの本文、図などが記載されている。このような文書は、図4に示すような木構造として表現することができる。すなわち、図3の文書は、序論全体、本論全体を示す2つの「節」が存在する。例えば、序論を示す「節」には、「表題」と「本文段落」が存在し、「表題」の内容として、「序論」という文字列が存在する。また、例えば、図は、本論全体を示す「節」に含まれ、さらに、「節」の中の「本文段落」に「図表領域」として含まれる。図は、図のイメージ部分と、図の表題により構成される。他の部分についても同様である。このように、構造化文書では、文書の構造を木構造で表わしている。

【0022】図5は、メモリ7に格納される構造化文書の構造の一例の説明図である。例示のための構造化文書9は、構造解析部1により文書の構造が解析され、例えば、図4に示す構造を有する構造化文書は、図5に示すようなデータ構造でメモリ7に格納される。すなわち、構造化文書の階層構造が、親子関係と兄弟関係にあるものの間の双方向にポインタを持つ木構造として記憶される。

【0023】構造化文書の木構造のノードとなる部分は、レコードで表現される。図5において、矩形で示し

6

ているそれぞれがレコードである。各レコードは、いくつかのフィールドを有する。図5では、レコードのIDとなる「No.」、レコードの種別を示す「属性」、テキストデータ等の実際に文書に現れる内容を示す「値」、検索式のために用いる情報を格納する「部分式」の各フィールドを示している。もちろん、このほかのフィールドを有していてもよいし、例えば、フィールド「部分式」は、このレコードとは別の種類のレコードで管理する構成であってもよい。

【0024】図4の「本論」という文字列を有する部分あるいは構造は、図5のフィールド「No.」として「1. 1. 1」を有するレコードに相当する。このレコードは、フィールド「属性」として「テキスト」を、フィールド「値」として文字列「本論」を有する。フィールド「No.」は、階層構造を表現するように構成されており、この値が「1. 1. 1」の場合には、値が「1」のレコードの子のレコード「1. 1」のさらに子であることを示している。上述のように、各レコードは、親、子、兄弟に対するポインタも有する。

【0025】図6は、知識ベース6に蓄積されている定型的な構造パターンの一例の説明図である。知識ベース6には、熟練したユーザが検索に用いるような、種々の定型的な部分構造のパターンが格納されている。例えば、図6のパターン1は、フィールド「属性」の値が「節」のレコードの子として、フィールド「属性」の値が「表題」のレコードと、フィールド「属性」の値が「本文段落」のレコードの2つのレコードを有し、「表題」のレコードの子としてどのような部分構造が接続されていてもかまわないという、部分構造を示している。パターン2も同様に、フィールド「属性」の値が「節」のレコードの子として「本文段落」のレコードが存在し、その子として「図表領域」のレコードが存在し、さらにその子として属性が未確定のレコードと、もう一つの子として「領域の表題」のレコードが存在し、「領域の表題」のレコードは子となるレコードを有する部分構造を示している。上述の属性が未確定のレコードの属性の確定方法については後述するが、これは知識ベースのルールとしては、どのような属性であっても良いことを表わしている。しかし、この属性は、検索式の合成の前までには、具体的に例示された構造の持つ属性に等しくなるように定められ、検索時には、その属性が条件として用いられる。

【0026】図4乃至図6を用いて、本発明の検索式作成支援システムの第1の実施例の具体的な動作を説明する。例えば、図4に示すような例示のための構造化文書9が構造解析部1で図5に示すような構造に展開されてメモリ7に格納され、また、図6に示すような定型的なパターンが知識ベース6に格納されているものとする。

【0027】ここで、図4に示す構造化文書の検索結果の例示として、ユーザが「領域の表題」を指示した場合

を考える。検索式構造抽出部2は、図6に示した知識ベース6の内容を参照し、指示された“領域の表題”を含む部分構造パターンとして、図6のパターン2を抽出する。このとき、複数の部分構造パターンが抽出されることがあるが、その場合には、抽出された複数の部分構造パターンを含むような部分構造パターンを生成したり、あるいは、ある基準に従って1つを選択するように構成することができる。

【0028】パターン2が抽出されたとき、パターン2と一致するレコードに対してマークが付与される。すなわち、図5に示した例では、太線で示したレコードにマークが付与される。マークを付与されたレコードは、メモリ7に展開されている構造の部分構造であり、検索式の持つべき構造となっている。

【0029】抽出されたパターン2には、“属性=?”や、“?”といった未確定のレコードが存在する。検索条件抽出部3では、知識ベース6から見つけ出した適用可能な定型的な部分構造パターンを確定するため、対応するレコードの内容を格納するフィールドから、必要な情報を抽出し、フィールド「部分式」に格納する。図6のパターン2の“属性=?”には、フィールド「No.」が「1. 2. 2. 1」であるレコードが対応する。このことから、この知識ベース6に格納されていたパターン2の構造に加え、“属性”の部分条件として用いるとよいということがわかる。そのため、フィールド「No.」が「1. 2. 2. 1」のレコードのフィールド「属性」の内容を、フィールド「部分式」に「属性=図」として設定する。

【0030】この例では存在しないが、マークが付与されたレコードの間にマークが付与されていないレコードが存在する場合には、そのレコードの数と、それに関する知識ベースの部分構造パターンとから、親子関係の構造の深さに関する自由度と、兄弟関係の間に存在してよい他の構造の数に関する自由度を決定し、フィールド「部分式」の一部として、これも格納する。

【0031】最後に、検索式合成部4により、図5で太線で示したマークが付与された部分構造のレコードを順次取り出し、検索を行なう装置が検索式として認識可能な形式の表現で検索式を合成する。図7は、生成された検索式の一例の説明図である。図7に示すように、検索式自体はわかりにくい表現であるが、本発明の検索式作成支援システムを用いることにより、検索式の表現方法を知らなくても、図7に示すような検索式が自動的に生成される。生成された検索式は、実際に検索を行なう装置が理解できる表現であればよい。ユーザは、合成された検索式を用いて実際に検索を行なう装置により検索を行なう。

【0032】図7を用いて具体的に示したように、一般に、構造化文書における検索式は木構造の形式で表現されるものであり、

1. 検索対象となるノードの属性の条件と、
2. 検索対象の周辺に現れるべきノードの属性の条件と、
3. 検索対象のノードと2. のノードの位置関係を特定するのである。例えば、図7に示した検索式では、検索対象は「領域の表題」であるから、「(属性=領域の表題)」という表現が検索対象となるノードの属性の条件であり、その他の「(属性=...)」という表現が検索対象の周辺に現われるべきノードの属性の条件であり、それらを結んでいる「/」や「#」等は位置関係を特定する条件である。位置関係を特定する条件においては、検索式中のノードとノードを結ぶ枝の、文書の木構造で実際にとる長さ(対応するノードとノード間に現れるノードの数)に関する制約条件が含まれる。

【0033】検索結果は、検索式中に記述された1. の属性の条件を満たす文書構造中のノードであって、2. に示したノード属性の条件を満たす文書構造中のノードが周辺に存在し、さらにそれが3. に記述された位置関係を満たすように存在するものである。すなわち、検索結果は、文書の木構造の中にある検索式と同形の部分構造(部分木)の中の1つのノードとなる。

【0034】次に、第1の実施例における検索式の編集に関する動作を説明する。上述のように、検索条件は、例示のために与えられた文書の構造の部分構造(部分木)となっている。一般に、検索条件を修正することは、この部分木の変形を行なうことに相当する。例えば、条件を強めようとする場合は、木の枝を追加したり、枝の長さを固定してしまえばよい。逆に、条件を弱める場合には、木の枝を刈ってしまったたり、枝の長さを自由に変えても良いようにすれば良い。

【0035】ところが、このように、条件式の部分木への枝の追加や枝刈りを自由に行なうと、例示した部分までもが検索できなくなってしまう危険がある。図8は、検索式の編集を行なう際の問題点の説明図である。図8の実線で示した部分は、図5においてマークが付与されて検索式となった部分木を抜き出した検索条件の木構造を示している。例えば、この木構造に、図8の破線で示した新たな枝(“属性=テキスト”の枝)を加えてしまうと、図5に示した構造化文書の木構造に重ねることができなくなる。そのため、検索式を生成する際に例示した部分構造が検索不能になってしまう。

【0036】文法的に正しいことの検査だけを行なった場合は、図8に破線で示す枝を追加した条件は正しいので、このような誤りを発見できない。そこで、本発明においては、編集時にも図5に示した構造を参照し、木構造への枝の追加や、長さの指定を行なうことに相当する編集作業を行なった場合には、その結果が図5の構造に重ねることができる場合のみを許すように構成している。このように、検索式編集部5は、誤り防止機能を含

【0037】図9は、検索式編集部5における編集操作の一例の説明図である。編集操作で行なえる変形は、例えば、図9に示したように、枝の追加、枝の削除、枝の長さの自由度、2つの枝の間（あるいは最も外側の枝のさらに外側）に存在して良い別の枝の数の自由度、及び、部分式のパターンの変更の5つの操作とすることができる。これらの編集操作に際して、検索式構造抽出部2、検索条件抽出部3等の機能を利用することもできる。

【0038】これらの操作の各々について、あるノードにおいて編集を行なうおうとした場合に、その操作が正しいか否かを次のようにして判定する。

【0039】まず、枝の追加の場合には、各ノードは図5に示したレコードに対応し、構造の親子関係と兄弟の関係はリンクされているので、枝の追加を指示された方向にこのリンクをたどり、レコードが存在すれば、枝の追加は正しいものと判定することができる。枝の削除の場合には、判定する必要がなく、いつでも正しい。ただし、削除する枝の両側の枝の間に存在して良い枝の数に関する自由度を無限大に設定する。

【0040】枝の長さの自由度の変更の場合、まず、自由度を大きくすることは、常に正しい。自由度を小さくするときは、図5の親子関係のリンクをたどり、その長さを調べ、設定された自由度がこの長さを含んでいれば正しいと判定する。

【0041】枝の間（最も外側の枝のさらに外側）に存在して良い枝の数の自由度の場合には、まず、自由度を大きくすることは、常に正しい。自由度を小さくするときは、図5の兄弟関係のリンクをたどり、その間の枝の数を調べ、指定された自由度がこの数を含んでいれば正しいと判定する。

【0042】部分式のパターンの変更の場合には、対応するレコードのフィールドを参照し、変更したパターンに整合するならば、正しいと判定する。

【0043】このように、検索式編集部5では、ユーザからの編集操作の指示に従い、検索式合成部4で生成された検索式に対して編集を行なうが、その際に、上述のような判定方法により正しいとされた操作のみを実行することによって、文法的にのみならず、意味的にも正しい式の変形のみを常に行なうことができる。

【0044】上述の第1の実施例に記載されている検索式作成支援システムを用いることによって、ユーザは検索結果の1つを例示するだけで検索式を作成することができる。これによって、構造化文書の内部構造や式の文法を知らない一般ユーザでも、検索式を簡単に得ることができるようになっていく。しかし、本質的に、例示されたものを検索可能な検索式は唯一ではなく、多数考えられる。このため、自動合成された検索式による検索結果に、意図したものに対する過不足が発生することは大いに有り得る。そのような場合には、検索式を修正を

なければならない。

【0045】上述の第1の実施例では、検索式の修正には、検索式編集手段を用いる。これにより、常に検索条件の正しさを保ちつつ、検索式の編集を行なうことができる。しかし、上述の検索式編集手段の実施例では、自動的に得られた検索式を修正したい場合に、検索式自体をユーザが修正することになる。このとき、どのように修正すれば良いかはユーザが自分で考えなければならない。すなわち、ユーザは、構造化文書の内部構造や式の文法に関するある程度の知識を再びここで要求されることになる。このように、検索式の修正は容易ではない。

【0046】以下に説明する第2の実施例では、誤った条件によって意図したものではない検索結果が得られた場合に、検索式側ではなく、検索結果の方の誤っている部分を指摘することにより、検索式を自動修正できる検索式作成支援システムについて説明する。

【0047】第1の実施例でも述べたように、一般に、検索式は木構造で表現されており、検索条件を修正することは、検索式の木構造の変形を行なうことに相当する。図9を用いて説明したように、例えば、図9の実線で示した部分木をもとの検索条件であるとする、これに図9の破線で示したような枝の状態に関する変化を加えることにより、検索条件を修正することができる。条件を強めようとする場合は、木の枝を追加したり、枝の長さを固定してしまえばよい。すなわち、もとの文書の検索結果の構造である木に重ねることができなくすれば、この部分は検索結果には含まれないことになる。逆に条件を弱める場合には、木の枝を削ってしまったら、枝の長さを自由に变えても良いようにすればよい。

【0048】図10は、本発明の検索式作成支援システムの第2の実施例を示すブロック図である。図中、21は検索式修正処理部、22は推論部、23は構造解析部、24はデータベース、25は構造化文書である。このほかに、図示しない入力部と検索部が存在する。

【0049】検索式修正処理部21は、ユーザからの誤りの指摘を入力部から受け取り、推論部22、構造解析部23を用いて、指摘された構造化文書の部分構造を検索結果として含めるように、あるいは、検索結果から除くように、検索式を修正する。

【0050】推論部22は、検索式修正処理部21からの要求に従い、検索式の修正の是非の判断や、検索式の修正案の提示、データベース24への登録と検索を行なう。データベース24には、推論部22への問い合わせに含まれている構造の各枝についての情報を格納している。各枝について、その枝が現在の検索式に含まれているか否か、その枝の長さの自由度はどのようになっているかに関する情報がすべて仮説として登録される。また、ユーザからの指摘等により正しくないといわれた仮説の組み合わせについて、正しくない旨の情報と、正しくないことの理由も登録される。正しくないことの理由

としては、ユーザが誤りとして指摘したノードを登録することができる。データベース24には、さらに、検索式の全体の木の大さき、取り除く枝の数、木の形等から、その枝に対する変更の是非を評価する評価関数が予め登録されている。評価関数は入力パラメタとして木の形を含んでいる。それゆえ、木の形と数式の対の形態で定義されている。

【0051】構造解析部23は、検索式修正処理部1の指示に従い、検索式の構造の解析を行なうとともに、構造化文書25の部分構造と検索式の構造を比較し、その比較結果を、あるいは、所定の部分構造を抽出する。例えば、推論部22から提示された修正案が、ユーザから指摘された枝を含むか否かを、構造化文書25の部分構造と修正案とを比較し、結果を報告する。また、追加すべきノードが存在する場合に、そのノードの周辺には存在せず、検索式にのみ存在する枝を抽出したり、削除するノードが存在するとき、その削除すべきノードの周辺には存在せず、正しいとわかっているノードの周辺には存在する枝を抽出する。

【0052】図11は、本発明の検索式作成支援システムの第2の実施例における動作の一例を説明するフローチャートである。検索式修正処理部21は、最初に、S31において上述の第1の実施例で自動合成した検索式、あるいは、ユーザが直接与えた検索式を現在の検索式として得る。そして、この検索式について、S32で検索式の木構造の各枝とその長さに関する状態を解析することを構造解析部23に依頼する。

【0053】S33において、構造解析部22からの返答に基づき、各枝の長さ、その枝が検索式に含まれることが正しいことを仮説として、推論部22に通知する。推論部22では、この通知に基づき、検索式の各枝についての情報をデータベース24へ登録する。

【0054】S34において、現在の検索式を用いて、図示しない検索部により検索を行ない、検索結果をユーザに提示する。ユーザは、提示された検索結果を参照し、S35において、ユーザが意図したものに対する過不足があるか否かを判断する。過不足があった場合には、S36において、ユーザは誤っている部分を順次指摘する。この指摘が検索結果の不足を指摘するものである場合には、S37からS38へ進み、指摘されたノードが検索結果に追加されるように、検索式を修正する処理を行なう。また、ユーザの指摘が冗長な検索結果を指摘するものである場合には、S37からS39へ進み、指摘されたノードを除外して検索されるように、検索式を修正する処理を行なう。そして、修正された検索式を現在の検索式として、S34へ戻って、更なる修正を行なうことが可能である。

【0055】図12は、追加ノード指定による検索式の修正処理の一例を示すフローチャートである。現在の検索式によって得られた検索結果に不足しているノードが

存在する場合について説明する。この場合、検索結果として含まなければならないものの例を、ユーザが指摘することになる。以下では、この指摘された部分を追加ノードと呼ぶことにする。指摘する追加ノードは、1つに限らず、複数指摘することが可能である。

【0056】検索式修正処理部21は、S41において、構造解析部23に対し、検索式の木構造と、構造化文書25における追加ノードおよびその周辺の構造を比較することにより、追加ノードの周辺には存在せず、検索式の構造にのみ存在するような、検索式の枝を全て見つけるように指示する。構造解析部23は、検索式の各枝ごとに、追加ノードの周辺の枝に、対応するものが存在するか否かを調べる。このとき、検索式の各枝は、枝の長さを自由に覚えて良いものとして調べ、検索式修正処理部21に対して返答する。枝の長さを固定された状態から自由に換えられる状態に変更して調べた場合には、枝の長さを自由に換えられる状態に変更した枝に関する情報も、返答に含まれる。

【0057】S42において、検索式修正処理部21は、推論部22に対し、前述の構造解析部23からの返答で得た、検索式の構造にのみ存在する枝を、全て検索式から取り除いて良いかどうかを問い合わせる。推論部22に問い合わせる際には、なぜそれらの枝を取り除きたいかという理由として、追加ノードを通知する。推論部22では、枝を取り除いてよいか否かを評価し、可否を検索式修正処理部21に通知する。同様に、S43において、構造解析部23で枝の長さを自由に換えられる状態に変更した枝に関しても、自由にして良いか否かを推論部22に問い合わせ、推論部22からの返答を受け取る。

【0058】S44において、推論部22から全ての許可が出ると、検索式の修正は可能である。このとき、推論部22は、許可を出すだけではなく、検索式への別の枝の追加や枝の長さの指定に関する指示をも出すので、これに従って、検索式修正処理部21は、検索式の木構造をその状態に変化させる。すなわち、S45で推論部22から許可の出た枝の削除を行ない、S46で推論部22から指示された枝の追加を行ない、S47で推論部22から指示された枝の長さの変更を行なう。これらの処理によって、検索式修正処理部21における追加ノードによる検索式の木構造の変更処理が完了する。

【0059】S44において、推論部22から1つでも許可が出なかった枝が存在する場合には、その部分は別の式で検索するべきであると判断し、修正を諦める。そして、ユーザにその旨を知らせたり、その部分のための検索式を自動合成する処理を呼び出すなどの措置を取る。

【0060】図13は、検索式の修正の具体例を説明するための構造化文書の部分構造の一例の説明図、図14は、修正対象となる検索式の構造の一例を示す説明図で

ある。例えば、図13に示すような構造を持つ構造化文書に対し、図14に示す破線部を含む構造を有する検索式で検索を行なう場合を考える。図中の*印は抽出位置を示している。この場合、検索結果は図13のノードa「序論は面白くなかったと思うが……」である。ここで、ユーザが図13のノードb「はじめに……」も検索結果として現れるべきであるという指摘を行なう。すると、構造解析部23によって、図14に示した検索式の構造と、図13のノードbを含む周囲の部分構造を比較し、ノードbの周囲には存在せず、検索式の構造にのみ存在する枝を抽出する。この例では、図14に破線で示した枝kが図13のノードbの周囲には存在しないので、構造解析部23によって、この枝kが抽出される。そして、推論部22がこの枝kを検索式の構造から取り除いて良いと許可すれば、検索式修正処理部21は検索式の構造から枝kを取り除き、図14に実線で示す構造のみとなるように検索式が修正される。そして、修正された検索式によって検索を行なうことにより、検索結果には、図13のノードaとノードbの両方の部分が含まれるようになる。

【0061】図15は、削除ノード指定による検索式の修正処理の一例を示すフローチャートである。検索結果として余分なものが得られた場合には、ユーザが検索結果として余分であるものの例を指摘することになる。以下では、この指摘された部分を削除ノードと呼ぶことにする。

【0062】ユーザが削除ノードを指示すると、S51において、検索式修正処理部21は推論部22に対し、現在の検索式が正しくないので修正方法の指示を出すように要求する。推論部22から修正の指示が来た場合は、S52からS53へ進み、検索式修正処理部21は、推論部22から受けた指示に基づく検索式の修正によって、削除ノードが検索結果に含まれないようになるかを調べるように構造解析部23に依頼する。S54において、構造解析部23から、削除ノードが含まれないようになるという返答を得た場合には、S55において、先に推論部22から得た修正の指示を採用し、その指示の通りに検索式の構造をその状態に変化させる。

【0063】S54において、推論部22から受けた修正指示にしたがっても依然として削除ノードが検索結果に含まれてしまう場合は、S56において、推論部22に対し、先の修正案が正しくないで新たな修正方法の指示を出すように要求する。そして、S52へ戻り、検索式修正処理部21は推論部22から正しい修正案が与えられるまでこれを繰り返す。

【0064】S52において、推論部22から修正できないとの返答を受けた場合には、S57以降の処理を行なう。S57において、検索式修正処理部21は構造解析部23に対し、すでに検索対象として正しいとわかっているノード周辺の構造と削除ノード周辺の構造に対し

て、検索式の木構造と重なる部分木に接している枝であって、正しいとわかっているノード周辺には存在し、なおかつ削除ノード周辺には対応する枝が存在しないような枝を全て見つけるように指示する。そして、S58において、構造解析部23によって抽出された枝のどれを検索式に追加したら良いかを推論部22に問い合わせる。このとき、追加したい理由として削除ノードを通知する。推論部22は、構造解析部23で抽出された枝の中から1本を選び、その長さを指定して検索式の木に追加するように、指示を検索式修正処理部21へ返す。推論部22から追加の指示が返された場合には、S59からS60へ進み、検索式修正処理部21は、推論部22からの追加の指示に従って、検索式の木構造をその状態に変化させる。

【0065】あるいは、S59において、推論部22がどの枝も追加しない方が良いと判断し、そのような返答が来た場合には、その部分は別の方法、例えば、ノードの属性などを変更して除外するべきであると判断し、修正を諦める。そして、ユーザにその旨を知らせたり、その部分を除外するためのノード属性を自動的に求める処理を呼び出すなどの措置をとる。

【0066】具体例として、上述の図13、図14を用いて説明する。例えば、図13に示すような構造を持つ構造化文書に対し、図14に示す実線部のみの検索式の構造により検索を行なった場合を考える。このときの検索結果は、図13のノードa「はじめに……」、および、ノードb「序論は面白くなかったと思うが……」の部分となる。ここで、ユーザが図13のノードbの部分は検索結果として現れるべきではないという指摘をする。と、検索式修正処理部21は推論部22に対して、修正方法の指示を要求する。ここでは、推論部22から修正できないとの返答を受けたとする。

【0067】検索式修正処理部21は構造解析部23に対して、図13のノードaとノードbの周囲の図14の実線部の木に相当する部分木を調べるように要求する。この例の場合には、正しいとわかっているノードaの検索式の木構造と重なる部分木に接する枝l, m, nを求める。このうち、枝mはノードbの周囲に対応する枝が存在するので、ここでは枝l, nが抽出される。抽出された枝l, nは、推論部22に送られ、このうちの1本が選択される。例えば、推論部22が枝lを検索式の枝に追加するのが最も良いと判断した場合には、検索式修正処理部21は、図14に破線で示すように、図13の枝lに相当する枝kを含むように検索式の構造を修正する。この修正によって、修正後の検索式を用いた検索結果には、図3のノードbの部分は含まれないようになる。

【0068】以上からわかるように、ユーザは誤っている部分をはじめから全て指摘する必要はなく、適当に指摘したならば、検索を実行し、この処理を繰り返し行な

い、意図した結果が得られたところで修正を止めれば良い。途中で検索の実行を行なうことは、本発明の検索式修正処理に何等影響を与えない。

【0069】図16ないし図19は、推論部22の動作の一例を説明するフローチャートである。推論部22に対する要求は、上述の説明からもわかるように、以下のようなものである。

要求1. 各枝が検索式に含まれるという情報と、その長さに関する情報とを仮説としてデータベース24に登録すること。(図11のS33)

要求2. 検索式から枝を取り除くか、長さを自由にする10 ことの是非を判断すること。(図12のS42, S43)

要求3. 現在の検索式が正しくない時に、修正案を提示すること。(図15のS51)

要求4. 検索式に加えることが可能な枝の候補の中から、1本の枝を選ぶこと。(図15のS58)

【0070】図16のS61において、推論部22は要求を受け付ける。S62において、受け付けた要求が上述の要求1ないし要求4のどの内容かを判定し、その要求に応じてS63ないしS66の処理のうちどれかを選択的に実行する。そして、S67において、要求元へ回答を返す。

【0071】以下、各々の要求に対する処理について、一例を説明する。上述の要求1に関しては、各枝が検索式に含まれるという情報と、その長さに関する情報とを仮説として、データベース24に登録を行なう処理のみを行なう。

【0072】要求2に関しては、図17のS71において、まず、検索式の部分木になっている木に対応する評価関数をデータベース24から検索する。ここで、部分木になっているか否かの判定は、構造解析部23が行なう。あらかじめ優先順位を考慮して木構造をデータベース24に格納しておくことにより、最初に検索されたデータを採用すればよいように構成することができる。検索された木構造に対応する評価関数から数式を取り出す。そして、S72において、その数式に木の形以外の引数、例えば、検索式の全体の大きさ、取り除く枝の数の情報を与えて計算を行なう。

【0073】S73において、計算された評価値が予め設定された範囲に入っているか否かを判定し、範囲に入っていない場合には、S74において、検索式から枝を取り除くこと、あるいは、長さを自由にすることを許可しないことを回答として、要求元へ返す。評価値が範囲に入っている場合には、S75において、その要求を許可する旨の回答を要求元へ返す。そして、S76において、現在の検索式に含まれる枝の状態の組み合わせが正しくないことを、データベース24に理由をつけて登録する。また、S77において、取り除くか、あるいは長さを自由にした枝に関して、データベース24内の情報 50

を、仮説が成立しない状態に更新する。

【0074】例えば、上述の図14に示した検索式の構造において、破線で示した枝kを検索式の枝から取り除く場合、図14に示す検索式の構造の部分木になっているデータをデータベース24から検索して取り出す。取り出されたデータと対になっている数式が、例えば、

$$(1 - (\text{削除する枝の数} / (\text{全体の木の枝の数})) * 100$$

という式であったとする。削除する枝の数は1本であり、全体の木の枝の数は5本である。これらの値を代入し、評価値は $(1 - 1/5) * 100 = 80$ となる。そして削除許可を出すのは、この値が60以上のときと定義されていれば、削除の許可を出すことになる。もちろん、削除許可を出すのを90以上と定義していれば、削除の許可は出されない。

【0075】ところで、この実施例では、以下でも2つの別の評価関数が利用される。それらについても実現方法は同様である。ただし、以下の2つの評価関数は、より早く正しい検索式へと到達するための工夫でしかないので、例えば、乱数や定数を返すものであっても良い。その場合は、ユーザが意図した式に到達するのに、少し多めに誤りの指摘が必要となるかもしれないということだけである。

【0076】要求3に関しては、まず、図18のS81において、現在の検索式に含まれる枝の状態の組み合わせが正しくないことをデータベース24に理由をつけて登録する。S82において、データベース24に登録されている仮説のどれかを翻し、新たな検索式の候補を求める。このとき、これまでの推論部22に対する要求によって、データベース24に登録されている、正しくないとかわっている枝の組み合わせを含んでしまうような状態以外の状態を全て求める。これにより、過去に正しくないと判断した履歴を用いた学習効果を得ることができる。求めた状態は複数存在する。そのため、S83において、各状態に関して用意された評価関数を用いて評価を行なう。S84において、計算した評価値に基づき、最も評価値の高いものを選ぶ。S85で評価値が所定の範囲内か否かを判定し、所定範囲外の場合には、S86で修正案がないことを回答として要求元へ返す。最高の評価値が所定範囲内の場合には、S87において、最高の評価値の状態を得るために翻した仮説の情報を修正案として要求元へ返す。そして、S88において、翻した仮説について、データベース24に反映させる。

【0077】要求4に関しては、図19のS91において、木の形等を基にデータベース24を検索して、追加の是非を評価するための評価関数を得る。そして、S92において、得られた評価関数を用いて、追加候補の各枝について、その枝と検索式の全体の木の大きさを入力として評価する。S93において、評価関数による評価結果の最も良いものを選ぶ。S94において、選択した

評価結果が有効なものか否かを判定し、有効でなければ S 9 5 において追加を許可しない旨を要求元へ返す。評価結果が有効であれば、S 9 6 において、選択した評価結果を有する枝を追加するように、要求元へ返答する。S 9 7 において、現在の検索式に含まれる枝の状態の組み合わせが正しくないことをデータベース 2 4 に理由をつけて登録する。また、S 9 8 において追加候補の各枝について、データベース 2 4 の情報を更新する。このとき、選択した枝については、その仮説が成立する状態として登録し、選択しなかった枝については仮説が成立しない状態として登録する。

【0078】上述のフローチャートでもわかるように、図 17 の S 7 6、図 18 の S 8 1、図 19 の S 9 7 等において、検索式に含まれる枝の集合が正しくないことをデータベースに登録し、図 18 の S 8 2 で、この登録されている正しくない枝の集合を含むような検索式の候補を排除している。これにより、2 度と同じ過ちを繰り返すことなく修正が行なわれていくので、指摘箇所が多くなれば、次第に確実に正しい式に近づけることができる。また、正しくない枝の状態の組み合わせには理由が 40 ついているので、この理由そのものを取り消した場合には、もとの状態に戻すことも可能であり、ユーザの誤り指摘の操作が誤りであった場合にも対処することができる。

【0079】この第 2 の実施例は、この構成を単独で用いることももちろん可能であり、その場合には、検索式が予め与えられることになる。また、第 1 の実施例における検索式編集部 5 として組み込むこともできる。この場合には、検索式構造抽出部 2、検索条件抽出部 3 等を用いて構造解析部 2 3 等の機能を実現することができる。また、データベース 2 4 を知識ベース 6 と共通とすることももちろん可能である。

【0080】

【発明の効果】以上の説明から明らかなように、本発明によれば、簡単な操作により自動的に検索式を作成することができ、例えば、構造化文書の内部構造や式の文法を知らないユーザでも、意図した式を簡単に得ることができる。これによって、一般ユーザが構造化文書の利点を有効に活用することが可能となる。また、習熟したユーザにも対しても有用であり、不注意による誤りがほぼ 40 完全に排除でき、意図した式を非常に短時間で正確に作成することが可能となる。また、検索式編集手段を有する構成では、常に検索条件の正しさを保ちつつ、検索式の編集を行なうことができるという効果がある。

【0081】また、検索結果の誤りを指摘するという簡単な操作によって、構造化文書の検索式を自動的に修正することができるので、構造化文書の検索式の文法を知らないユーザでも、意図した式を簡単に得ることができる。これによって、一般のユーザが構造化文書の利点を有効に活用することが可能となる。また、習熟したユー

ザに対しても有用であり、不注意による誤りがほぼ完全に排除でき、意図した式を非常に短時間で正確に作成することが可能となるという効果がある。

【図面の簡単な説明】

【図 1】 本発明の検索式作成支援システムの第 1 の実施例の構成図である。

【図 2】 本発明の検索式作成支援システムの第 1 の実施例における検索式生成の動作を示すフローチャートである。

【図 3】 構造化文書の一例の説明図である。

【図 4】 構造化文書の構造化された内部表現の一例の説明図である。

【図 5】 メモリ 7 に格納される構造化文書の構造の一例の説明図である。

【図 6】 知識ベース 6 に蓄積されている定型的な構造パターンの一例の説明図である。

【図 7】 生成された検索式の一例の説明図である。

【図 8】 検索式の編集を行なう際の問題点の説明図である。

【図 9】 検索式編集部 5 における編集操作の一例の説明図である。

【図 10】 本発明の検索式作成支援システムの第 2 の実施例を示すブロック図である。

【図 11】 本発明の検索式作成支援システムの第 2 の実施例における動作の一例を説明するフローチャートである。

【図 12】 追加ノード指定による検索式の修正処理の一例を示すフローチャートである。

【図 13】 検索式の修正の具体例を説明するための構造化文書の部分構造の一例の説明図である。

【図 14】 修正対象となる検索式の構造の一例を示す説明図である。

【図 15】 削除ノード指定による検索式の修正処理の一例を示すフローチャートである。

【図 16】 推論部 2 2 の全体動作の一例を説明するフローチャートである。

【図 17】 推論部 2 2 における検索式から枝を取り除くか、長さを自由にするかのは非を判断する動作の一例を説明するフローチャートである。

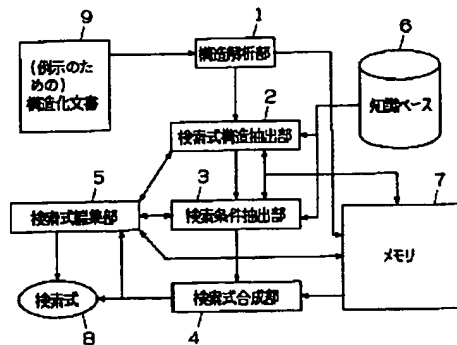
【図 18】 推論部 2 2 における修正案を提示する動作の一例を説明するフローチャートである。

【図 19】 推論部 2 2 における追加候補の枝の中から 1 本を選ぶ動作の一例を説明するフローチャートである。

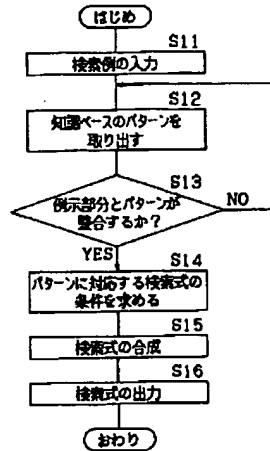
【符号の説明】

1…構造解析部、2…検索式構造抽出部、3…検索条件抽出部、4…検索式合成部、5…検索式編集部、6…知識ベース、7…メモリ、8…検索式、9…構造化文書、21…検索式修正処理部、22…推論部、23…構造解析部、24…データベース、25…構造化文書。

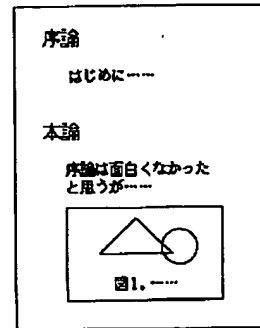
【図1】



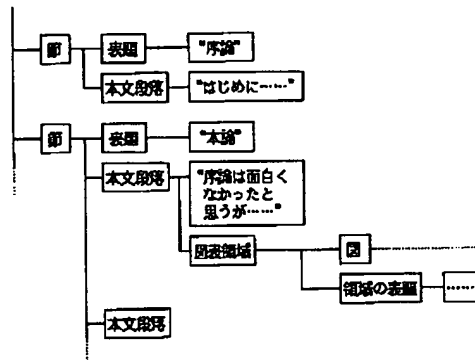
【図2】



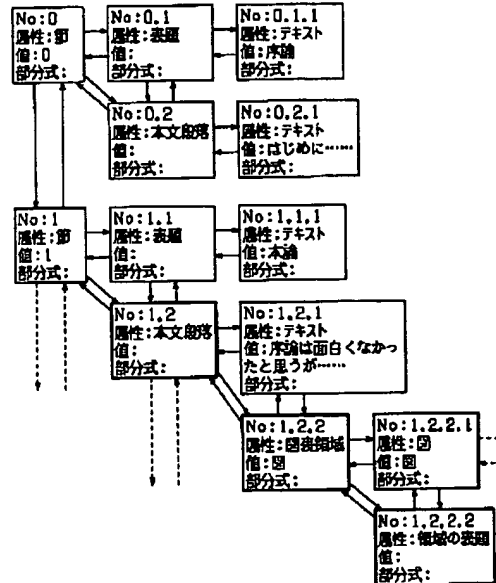
【図3】



【図4】



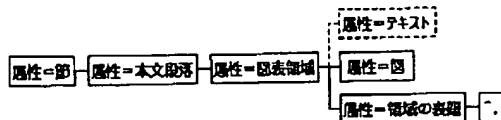
【図5】



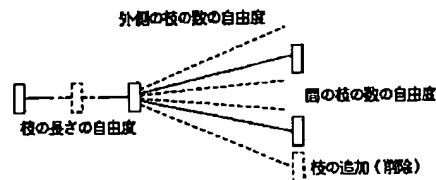
【図7】

(属性=節)/(属性=本文段落)/(属性=図表領域)/
((属性=図)((属性=領域の表題)/.))

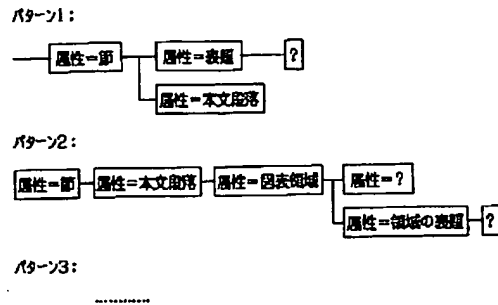
【図8】



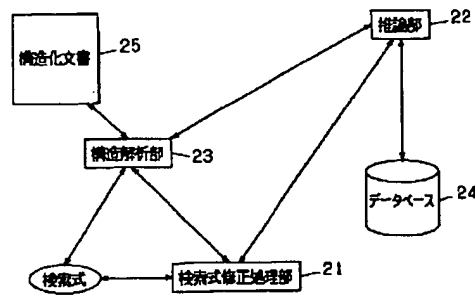
【図9】



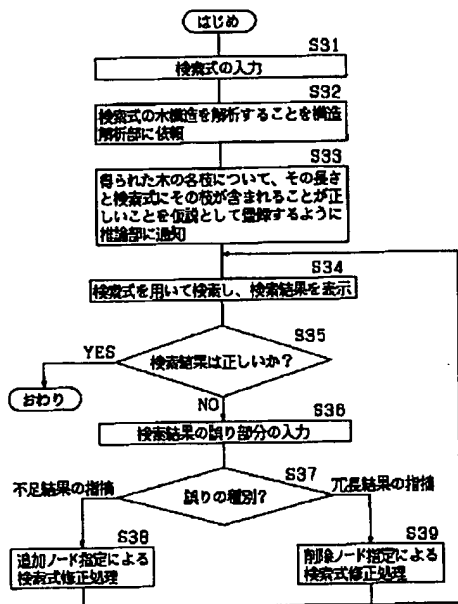
【図6】



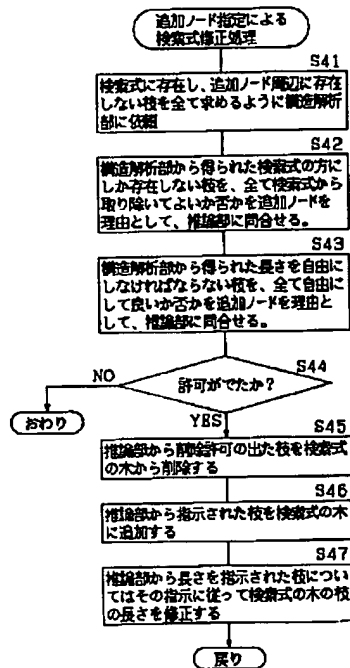
【図10】



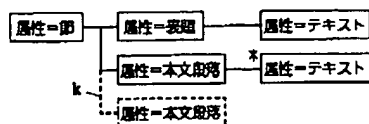
【図11】



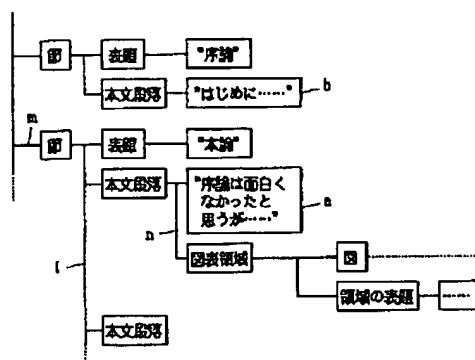
【図12】



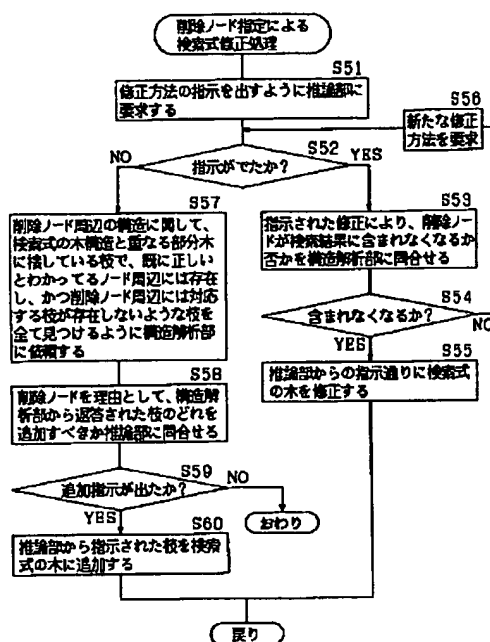
【図14】



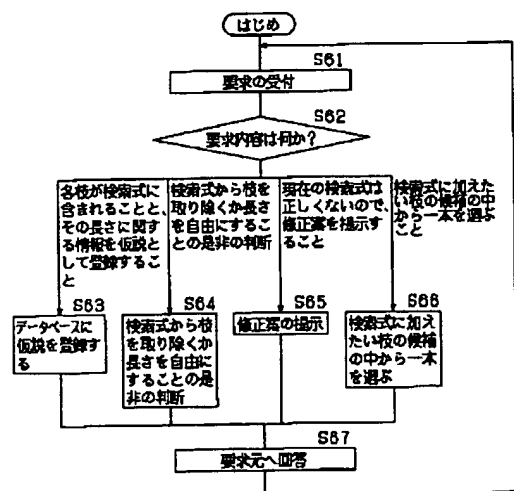
【図13】



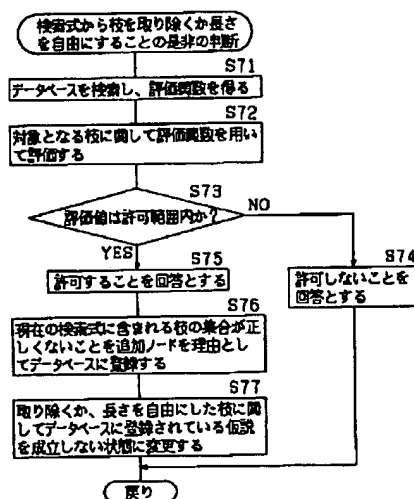
【図15】



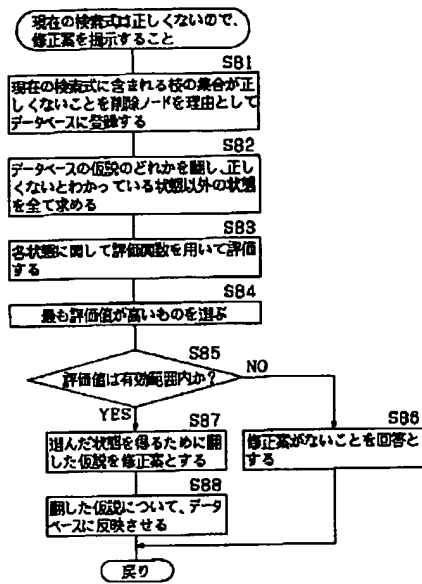
【図16】



【図17】



【図18】



【図19】

